824

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/595,719	06/16/2000	Minos N. Garofalakis	Garofalakis-6-1-36-11-10	1456

7590 12/15/2004

Stephen J Weed Esquire
Synnestvedt & Lechner LLP
2600 Aramark tower
1101 Market Street
Philadelphia, PA 19107-2950

EXAMINER	
HUYNH, CONG LAC T	
ART UNIT	**PAPER NUMBER**
2178	

DATE MAILED: 12/15/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/595,719	GAROFALAKIS ET AL.	
	Examiner	Art Unit	
	Cong-Lac Huynh	2178	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *09 June 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-22* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,2,7,8 and 13-22* is/are rejected.

7)☒ Claim(s) *3-6 and 9-12* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This action is responsive to communications: amendment filed 6/9/04 to the application filed on 6/16/00.

2. Claims 1-22 are pending in the case. Claims 1, 7, 14, 16, 18 are independent claims.

3. The rejections of claims 1-2, 7-8, 13-22 under 35 U.S.C. 103 103(a) as being unpatentable over Tateno have been withdrawn in view of Applicants arguments.

Claim Objections

4. Claim 18 is objected to because of the informalities. As seen in the claim, it appears that step "generalizing input sequences ..." includes "discovering OR patterns ...", "discovering sequence patterns ..." and "selecting a document descriptor ..." This is not consistent with the specification (pages 10-14) where step "generalizing input sequences ..." includes only "discovering OR patterns ..." and "discovering sequence patterns ...", and does not include step "selecting a document descriptor ...." If it is a typographical error, please reset the indents to show the inclusion of the two discovering steps within the generating step.

Claim Rejections - 35 USC § 102

5. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –

> (e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

The changes made to 35 U.S.C. 102(e) by the American Inventors Protection Act of 1999 (AIPA) and the Intellectual Property and High Technology Technical Amendments Act of 2002 do not apply when the reference is a U.S. patent resulting directly or indirectly from an international application filed before November 29, 2000. Therefore, the prior art date of the reference is determined under 35 U.S.C. 102(e) prior to the amendment by the AIPA (pre-AIPA 35 U.S.C. 102(e)).

6. Claims 7-8 are rejected under 35 U.S.C. 102(e) as being anticipated by Papakonstantinou et al, DTD Inference for views of XML data, ACM May 2000, pages 35-46.

Regarding independent claim 7, Papakonstatinou discloses:

- generalizing input sequences associated with a document to develop general sequences, said input sequences reflecting the structure of a document (**pages 35-36**: ".. XML marks the 'return of the schema' (albeit loose and flexible) in semistructured data, in the form of its Data Type Definition (DTDs) ... DTDs describes the structure of the objects (elements) participating in an XML document ...", "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees ... the generalization to multiple sources is straightforward, since these can be viewed as one source ....")

- selecting a document descriptor from said input sequences, said general sequences where said factored sequences using minimum descriptor length (MDL) principles (**pages 35-36**: "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees to which one of the variables in the tree pattern binds ... constructing a tight ltd for the view, i.e. an ltd that precisely characterizes the type structures of trees ... we overcome these limitations by enhancing ltds with a simple subtypying mechanism ...specialized ltds encompass the expressive power of formalism.. "; the fact that the specialized ltds are *simple*, *precisely characterizes the type structure of trees* and encompass the data structure implies that the ltds, which is the DTDs, are selected from the tag sequence using the minimum descriptor length; **page 36**: "... the 'pattern' ... the limited from of inference can be accomplished by inferring the pattern that view variables may bind to ...")

Regarding claim 8, which is dependent on claim 7, Papakonstatinou discloses:

- encoding said input sequences, said general sequences, and said factored sequences (pages 35-38: the tags of the document, which are the sequences, are encoded data)
- selecting a document descriptor which encompasses all of said input sequences and exhibits a minimum MDL cost (**pages 35-36**: "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees to which one of the variables in the tree pattern binds ... constructing a tight ltd for the view, i.e.

an ltd that precisely characterizes the type structures of trees ... we overcome these limitations by enhancing ltds with a simple subtypying mechanism ...specialized *ltds encompass* the expressive power of formalism.. "; the fact that the specialized ltds are *simple*, *precisely characterizes the type structure of trees* and encompass the data structure implies that the ltds, which is the DTDs, are selected from the tag sequence using the minimum descriptor length, and thus have the minimum cost)

7. Claims 16-19 are rejected under 35 U.S.C. 102(e) as being anticipated by Moh et al., Re-engineering Structures from Web Documents, ACM June 2, 2000, pages 67-76.

Regarding independent claim 16, Moh discloses:

- discovering OR patterns among said input sequences (pages 69, 73-74)
- discovering sequence patterns among said input sequences and OR patterns (pages 74-75)

Regarding claim 17, which is dependent on claim 16, Moh discloses that discovering OR patterns comprises partitioning said input sequences (page 73).

Regarding independent claim 18, Moh discloses:

Generalizing input sequences, said generalizing comprises:

- discovering OR patterns among said input sequences (pages 69, 74)

- discovering sequence patterns among said input sequences and OR patterns (pages 74-75)

Selecting a document descriptor from said input sequence and said general sequences (page 72, Final Construction of DTD).

Regarding claim 19, which is dependent on claim 18, Moh discloses that discovering OR patterns comprises partitioning said input sequences (page 73).

Claims 14-15 are for a computer readable medium of method claims 16-17, 18-19, and are rejected under the same rationale.

Claim Rejections - 35 USC § 103

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to

consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

10. Claims 1-2 are rejected under 35 U.S.C. 103(a) as being unpatentable over Papakonstatinou et al., DTD Inference for Views of XML Data, ACM May 2000, pages 35-46.

Regarding independent claim 1, Papakonstatinou discloses:

- generalizing input sequences associated with a document to develop general sequences, said input sequences reflecting the structure of a document (**pages 35-36**: ".. XML marks the 'return of the schema' (albeit loose and flexible) in semistructured data, in the form of its Data Type Definition (DTDs) ... DTDs describes the structure of the objects (elements) participating in an XML document ...", "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees ... the generalization to multiple sources is straightforward, since these can be viewed as one source ....")
- selecting a document descriptor from said input sequences, said general sequences where said factored sequences using minimum descriptor length (MDL) principles (**pages 35-36**: "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees to which one of the variables in the tree pattern binds ... constructing a tight ltd for the view, i.e. an ltd that precisely characterizes the type structures of trees ... we overcome these

limitations by enhancing ltds with a simple subtypying mechanism ...specialized ltds encompass the expressive power of formalism.. "; the fact that the specialized ltds are *simple*, *precisely characterizes* *the type structure of trees* and encompass the data structure implies that the ltds, which is the DTDs, are selected from the tag sequence using the minimum descriptor length; **page 36**: "... the 'pattern' ... the limited from of inference can be accomplished by *inferring the pattern* that view variables may bind to ..")

- grouping the tags where the tags showing the input sequence of the structure of the document (pages 37-39, Example 2.2, Example 2.7, Example 2.13)

Papakonstatinou does not explicitly disclose factoring said input sequences and said general sequences to develop factored sequences.

However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Papakonstatinou to include factoring the input sequence and said general sequences to develop factored sequences for the following reason. The grouping of tag sequences in Papakonstatinou, suggests that the tag sequence of a document in Papakonstatinou, while simple and encompass the formalism of data suggests that the tag names of the same types are grouped together for a precise DTD with a shortest length.

Regarding claim 2, which is dependent on claim 1, Papakonstatinou discloses:

- encoding said input sequences, said general sequences, and said factored sequences (pages 35-38: the tags of the document, which are the sequences, are encoded data)
- selecting a document descriptor which encompasses all of said input sequences and exhibits a minimum MDL cost (**pages 35-36**: "... variable bindings extracted by the tree pattern ... extract from the input the list of subtrees to which one of the variables in the tree pattern binds ... constructing a tight ltd for the view, i.e. an ltd that precisely characterizes the type structures of trees ... we overcome these limitations by enhancing ltds with a simple subtypying mechanism ...specialized *ltds encompass* the expressive power of formalism.. "; the fact that the specialized ltds are *simple*, *precisely characterizes the type structure of trees* and encompass the data structure implies that the ltds, which is the DTDs, are selected from the tag sequence using the minimum descriptor length, and thus have the minimum cost)

11. Claims 1-2, 7-8, 13, 20-22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Moh et al. Re-engineering Structures from Web Documents, ACM, June 2, 2000, pages 67-76.

Regarding independent claim 1, Moh discloses:

- generalizing input sequences associated with a document to develop general sequences, said input sequences reflecting the structure of a document (page 74: the sequence of a document is generalized)
- selecting a document descriptor from said input sequences, said general sequences, and said factored sequences using minimum descriptor length (MDL) principles (pages 74-76: the document DTD is derived from the sequence of document elements to reduce the repeated elements and thus providing a DTD with minimum descriptor length)

Moh does not disclose factoring said input sequences and said general sequences to develop factored sequences.

However, Moh does teach structural clustering of document tags (page 69).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Moh to include the factoring step into Moh since clustering the structure of a document via the input sequence of the document tags in Moh suggests the repeated tags be clustered, which means be grouped together to form a short sequence. It was well known in the art to cluster the repeated items such as the same elements in a web page, or documents of the same topic to form a collection. The combination of factoring step into Moh would help to accurately derive a precise DTD for a document collection.

Regarding claim 2, which is dependent on claim 1, Moh discloses:

- encoding said input sequences, said general sequences, and said factored sequences (pages 74-75: document tags are encoded)
- selecting a document descriptor which encompasses all of said input sequence, and exhibits a minimum MDL cost (pages 74-76)

Claims 7-8 include the limitations of claims 1-2, and are rejected under the same rationale.

Regarding claim 13, which is dependent on claim 7, Moh does not disclose explicitly that factoring said input sequences and said general sequences to develop factored sequences, wherein said factored sequences are available to said selecting. However, Moh does teach structural clustering of document tags (page 69). It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Moh to include the factoring step into Moh since clustering the structure of a document via the input sequence of the document tags in Moh suggests the repeated tags be clustered, which means be grouped together to form a short sequence. It was well known in the art to cluster the repeated items such as the same elements in a web page, or documents of the same topic to form a collection. Therefore, the combination of factoring step into Moh would help to accurately derive a concise and precise DTD for a document collection.

Regarding claim 20, which is dependent on claim 19, Moh does not disclose explicitly that factoring said input sequences and said general sequences to develop factored sequences, wherein said factored sequences are available to said selecting. However, Moh does teach structural clustering of document tags (page 69). It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Moh to include the factoring step into Moh since clustering the structure of a document via the input sequence of the document tags in Moh suggests the repeated tags be clustered, which means be grouped together to form a short sequence. It was well known in the art to cluster the repeated items such as the same elements in a web page, or documents of the same topic to form a collection. Therefore, the combination of factoring step into Moh would help to accurately derive a concise and precise DTD for a document collection.

Regarding claim 21, which is dependent on claim 20, Moh does not disclose explicitly that said selecting employs minimum descriptor length (MDL) principles (page 71).

Regarding claim 22, which is dependent on claim 21, Moh discloses that said document descriptor is a document type descriptor (DTD) and said document is an extensible Markup Language (XML) document (pages 67-75).

Allowable Subject Matter

12. Claims 3-6, 9-12 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

13. Applicant's arguments with respect to claims 1 and 16 have been considered but are moot in view of the new ground(s) of rejection.

Regarding independent claim 1, Applicants argue that Tateno does not disclose or suggest the features of "factoring" and "minimum descriptor length principles" since it is unclear how these features are applicable to an electronic document containing a well-defined structure in Tateno (Remarks, page 7).

Examiner agrees.

Papakonstantinou and Moh suggest the argued feature (see the rejection above).

Regarding independent 16, Applicants argue that Tateno fails to teach or suggest discovering sequence patterns among input sequences and OR patterns, as these are outside the scope of his analysis of the defined DTD structure (Remarks, page 8).

Moh discloses the argued feature (see the rejection above).

Conclusion

14. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Chen et al. (US Pat No. 6,766,330 B1, 7/20/04, filed 10/12/00).

Sundaresan et al. (US Pat No. 6,651,059, 11/18/03, filed 11/15/99).

Barrett (US Pat No. 6,134,512, 10/17/00, filed 11/24/97).

Leppinen et al. (US Pat No. 6,675,219 B1, 1/6/04, filed 11/1/99).

Strong (US Pat No. 6,167,523, 12/26/00, filed 5/5/97).

Murashita (US Pat No. 6,330,574 B1, 12/11/01, filed 3/30/98).

GAJRAJ (US Pat App Pub No. 2002/0002566 A1, 1/3/02, filed 7/16/98).

Kougiouris et al. (US Pat App Pub No. 2004/0039993 A1, 2/26/04, filed 8/27/03, priority 11/12/99).

Perycz et al. (US Pat App Pub No. 2003/0056193 A1, 3/20/03, filed 9/17/01).

Lennon (US Pat App Pub No. 2003/0208473 A1, 11/6/03, filed 1/28/00).

Royal (US Pat App Pub No. 2001/0027459 A1, 10/4/01, filed 2/28/01, priority 3/1/00).

Fong et al. (US Pat App Pub No. 2002/0085032 A1, 7/4/02, filed 7/6/01, priority 12/23/97).

Dodge, Using SGML to Streamline Print and CD-ROM Production, CD-ROM Professional, Mar 1994, vol. 7, iss. 2, pg. 77, 5 pgs.

Ashish et al., Wrapper Generation for Semi-Structured Internet Sources, ACM December 1997, pages 8-15.

Wallace et al., Haskell and XML: Generic Combinators or Type-Based Translation?, ACM 1999, pages 148-159.

Bergamaschi et al., An Approach for the Extraction of Information from Heterogeneous Sources of Textual Data, Google August 1997, pages 1-7.

Poulin et al., The Oher Formalization of Law: SGML Modelling and Tagging, ACM 1997, pages 82-88.

Adelberg, NoDoSE—a Tool for Semi-Automatically Extracting Structured and Semistructured Data from Text Documents, ACM June 1998, pages 283-294.

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Cong-Lac Huynh whose telephone number is 571-272-4125. The examiner can normally be reached on Mon-Fri (8:30-6:00).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Stephen Hong can be reached on 571-272-4124. The fax phone number for the organization where this application or proceeding is assigned is 571-273-4125.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Cong-Lac Huynh
Examiner
Art Unit 2178
12/01/04